Weekly Report

November 18, 2017

1 Work

This week, we have employed k-means method to speed up the embedding process of Largevis. We have achieved about 5 times acceleration on MNIST. The whole embedding process includes three stages:

1) At first, all data points are initialized randomly. Similar data will group into small cluster during the optimization process.

2) Since we have observed many small clusters, it is more efficient to merge two similar clusters by moving the whole cluster instead of moving each point. Therefore, we perform k-means to group small clusters. When we compute the gradient for a data object, we assign the gradient to the whole cluster which the data object belongs to.

3) Similar clusters will merge into a large cluster within several steps. Then, we stop using k-means result and refine the embedding result at the data point level.

1.1 工作进度

Table 1: 工作进度

TASK	PROGRESS	DATE
dimension reduction	We plan to add product quantization to speed up knn graph construction andevaluation module to do experiments.	11.30
*2Vec survey	Write a little framework	12.30

2 Paper Reading

2.1 Learning Transductive Network Embedding

文章提出了一个半监督网络嵌入方法，在网络嵌入的过程中编码一部分样本标签信息，从而在最终做预测或者分类任务的时候提升效果。

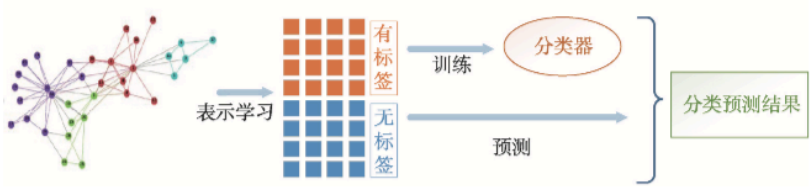

Fig.1 Traditional unsupervised embedding learning and classification

图1 传统无监督表示学习和分类

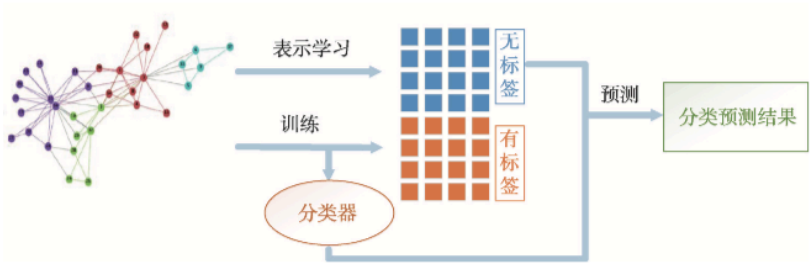

Fig.2 Transductive embedding learning and classification

图2 直推式表示学习和分类

Figure 1: 1

2.2 Fast, Warped Graph Embedding: Unifying Framework and One-Click Algorithm

文章是总结了目前网络学习的整体框架，大致分为三个部分：1）节点的原始相似性度量；2）节点嵌入之后的相似性度量；3）两个相似性的误差度量。最终的目的是调整节点的嵌入结果，使得误差度量最小。

2.3 ARC: A Pipeline Approach Enabling Large-Scale Graph Visualization

文章提出了一个图布局的方法，其中考虑到节点大小和布局的紧凑性。

2.4 Asymmetric Transitivity Preserving Graph Embedding

HOPE是一个图嵌入方法，对于有向图中的每一个节点学习了两个向量：出向量和入向量，这样能够更好地刻画有向图的结构。

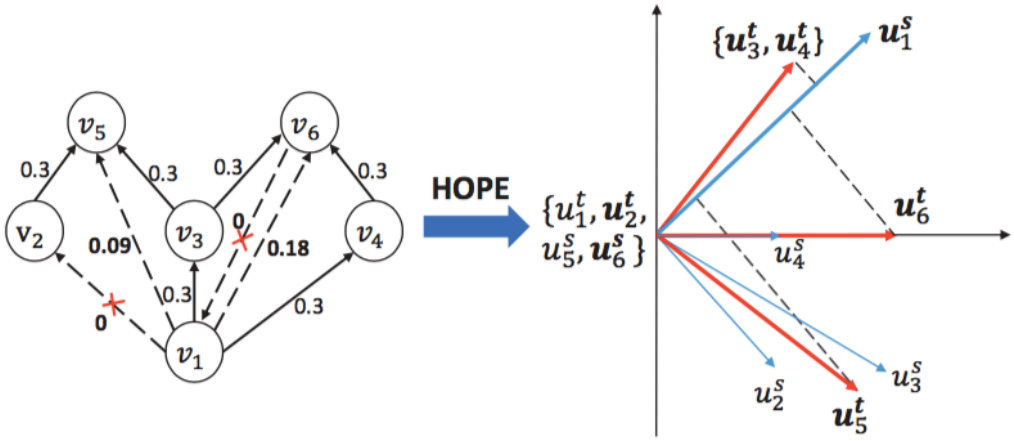

Figure 2: 4

2.5 Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization

文章使用拉格朗日乘子法，将原始问题转换为对偶空间上的问题进行求解。结果证明，使用对偶空间上的轴上升方法会比原始空间随机梯度更快得到高精度解。